An Investigation of the Accuracy of Heuristic Methods for Cost Uncertainty Analysis

Major Wendell P. Simpson III, Ph.D
Major Kevin P. Grant, Ph.D

Air Force Institute of Technology/LAS
2950 P Street
Wright Patterson AFB, OH 45433-7765

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

AU-AFIT/LA-TR-94-1

An Investigation of the Accuracy of Heuristic Methods for Cost Uncertainty Analysis

Major Wendell P. Simpson III, Ph.D
Major Kevin P. Grant, Ph.D

Air Force Institute of Technology/LAS
2950 P Street
Wright Patterson AFB, OH 43433-7765

9

An Investigation of the Accuracy of Heuristic Methods for Cost Uncertainty Analysis

Major Wendell P. Simpson III, Ph.D
Major Kevin P. Grant, Ph.D

ABSTRACT

Cost uncertainty analysis has received a significant amount of attention over the last several years. The purpose of a cost uncertainty analysis is to identify the cost and schedule implications associated with program uncertainties. Common methods for uncertainty analysis characterize the possible cost and schedule outcomes of a project using a probability density function (pdf). Heuristic methods have been proposed for uncertainty analysis that assume the shape of the total cost pdf is either normally or lognormally distributed. While experienced analysts feel these distributions provide reasonable approximations, little evidence exists to either confirm or refute these presumptions. This research examines the accuracy of the heuristic methods under varying conditions. An experiment is conducted in which the number of cost elements, the degree of skewness of the cost element distributions, and the degree of correlation between cost elements are systematically varied. The resulting total cost distributions are compared to the heuristic distributions using goodness of fit tests. The results show that the normal distribution provides an excellent approximation for the simulated distribution. Guidelines are offered that help the cost analyst determine whether these heuristics ought to be applied in a cost uncertainty analysis.

An Investigation of the Accuracy of Heuristic Methods for Cost Uncertainty Analysis

1. Introduction

Declining budgets and technological advances have fueled the recent demand for cost uncertainty analyses. During the Reagan defense buildup, uncertainty analyses were often treated superficially or neglected by the service departments and OSD. Adding an ECO/Risk line to the estimate and applying a factor was usually adequate enough to satisfy the requirement for an uncertainty analysis. Today, there is an expectation in DOD that uncertainty analyses will be far more comprehensive and that they will accurately characterize the cost and schedule implications associated with program uncertainties. The cost analysis community has responded to this challenge to deliver a credible, comprehensive uncertainty analysis. There has been a significant increase over the last several years in both research and training programs addressing cost uncertainty analyses.

The heightened interest in uncertainty analyses should be welcomed by cost analysts. Most cost analysts are understandably uncomfortable with providing a single estimate of the cost of a system. Only one thing can be said of the point estimate with complete certainty–it will be wrong! A cost uncertainty analysis provides the analyst with an opportunity to quantify the hesitations, the qualifications, and the hems and haws that accompany most inputs received in the data collection process. Attaching a range and probabilities to input values allows analysts to formally characterize and communicate the uncertainty inherent in the inevitable "soft" inputs used in a cost analysis.

The product of a statistical uncertainty analysis is a probability density function (pdf) of total system cost. An example of a total cost pdf is shown in Figure 1. The pdf depicts the probabilities associated with possible values for total system cost. Not all mathematical functions are pdfs–the term is reserved for functions that have certain properties. One such property stipulates that the area under the curve of a proper pdf will be equal to one.

Insert Figure 1 Here

The total cost pdf provides analysts with two valuable pieces of information. First, the pdf gives the analyst an estimate of the range in which total system cost is likely to fall. Second, the total cost pdf provides an estimate of probabilities associated with possible values of total system cost. This second piece of information stems from the property mentioned above. To estimate the probability that cost will not exceed some value, x_1, the analyst need only calculate the area under the total cost pdf between the lowest possible value for cost and x_1. Transforming the total cost pdf into a cumulative density function (cdf) aids in making these calculations. Using the total cost cdf, the analyst can answer this same question by simply locating x_1 on the x-axis, reading up to the cdf and over to the y-axis to estimate probabilities. Transforming a pdf into a cdf is a straightforward exercise of accumulating probabilities (see Figure 1). As such, the goal of the cost uncertainty analysis remains generating the total cost pdf.

A statistical uncertainty analysis generally begins by specifying pdfs for the uncertain inputs associated with the cost estimate. This paper will focus on additive uncertainty models in which the input pdfs usually describe the uncertainty associated with lower level cost elements[1]. The total cost pdf is then constructed by accumulating (summing) the input pdfs. There are a number of tools available to help the analyst generate the total cost pdf from the input distributions. One way to classify these tools would be to characterize the methods as either simulation or heuristic techniques. The

[1] An *additive* uncertainty model is distinguished from a *multiplicative* uncertainty model by the way in which the component pdfs are combined. An additive model is one in which the total cost pdf is the result of *summing* lower level pdfs. A multiplicative model is one in which two or more pdfs are *multiplied* together in the process of arriving at the total cost pdf. For example, a multiplicative model might contain a pdf to describe manufacturing hours and another pdf for the manufacturing wrap rate. In the process of arriving at the total cost pdf, these two pdfs would have to be multiplied together.

difference emphasized by this classification is the manner in which the shape of the total cost pdf is obtained. Monte Carlo simulation derives the shape of the total cost pdf through repeated random sampling from the cost element pdfs. Heuristic methods arrive at the total cost pdf by assumption. An example of a heuristic method is Young's FRISK model (1992). The FRISK model simply assumes that total cost can be accurately represented by the lognormal distribution.

Both Monte Carlo simulation and heuristic methods generate approximations of the total cost pdf. Analysts are forced to accept an approximation because the "true" total cost pdf cannot be determined for problems of the size normally encountered. The "true" total cost pdf can be derived with analytic techniques, but these techniques can only be successfully applied when there are very few cost elements. Monte Carlo simulation has been established as an accurate method for cost uncertainty analysis[2] (Deinemann 1966). The technique is accurate in that the total cost pdf produced by a simulation is a close approximation of the "true" total cost distribution (Burgess and Book 1993). Heuristics make no attempt to derive the total cost pdf–a shape is simply assumed. Heuristics apply a predetermined shape to the total cost pdf regardless of the nature of the pdfs for the component cost elements. The accuracy of the heuristic approaches depends then on the characteristics of the component cost element pdfs.

The obvious question becomes–why assume a shape for the total cost pdf when you can derive an accurate representation using simulation? The primary reason for selecting a heuristic is speed of computation. Monte Carlo simulation is accurate, but the accuracy can be expensive. The price paid is the time required to execute a Monte Carlo simulation. Even with a fast PC executing a state-of-the-art simulation program, a single run of a simulation with 50 cost elements can take well over an hour to execute (Graham 1992). If a heuristic method is applied, a single run of a simulation model is replaced by a single, near-instantaneous calculation. It is reasonable when conducting an uncertainty analysis to expect that many such runs of the uncertainty model will be

[2] The accuracy of Monte Carlo simulation improves with the number of sampling iterations performed. The analyst controls the number of iterations in a Monte Carlo simulation.

required. Revised input data, error corrections, and sensitivity analyses can result in the requirement for dozens of runs. Over the course of an uncertainty analysis, a significant time savings can be realized by applying heuristic techniques rather than simulation.

When selecting a methodology for a statistical cost uncertainty analysis, the analyst must evaluate the trade-off that exists between the speed of the heuristic methods and the accuracy of the simulation. Most cost analysts would sacrifice accuracy for speed only under duress. Unfortunately, time duress is a often a reality in the cost analysis business. Currently, little or no information exists to aid in this decision. How much accuracy is sacrificed when a heuristic approach is adopted? How is accuracy of a heuristic method affected by the nature of the cost element pdfs? The purpose of this paper is to communicate the results of research directed at providing this information. The goal is to provide guidelines that analysts can use to judge the accuracy of heuristics for their particular application.

The following section of the paper is dedicated to a review of some fundamentals of cost uncertainty analysis. Section 3 provides an outline of the experiment conducted while section 4 provides the results. The final section is reserved for conclusions and recommendations.

2. Some Fundamentals

In order to fully describe a total cost pdf, three separate pieces of information are needed: the location (mean) of the pdf, the dispersion (variance) of the pdf, and the shape of the pdf. Given an additive uncertainty model, the first two pieces of information can be determined directly given only pencil and paper. This is true regardless of whether simulation or heuristic methods are selected for the uncertainty analysis. However, knowing the mean and variance of total cost does not completely specify the total cost pdf. It is possible for two pdfs to have the same mean and variance, and still have different shapes. This is where the choice of uncertainty method comes to bear.

Assume that an additive model is being used such that total cost is equal to the sum of n WBS elements. Assume that the n WBS elements are represented by a pdf that is fully specified. In other words, the pdfs for each of the i component elements have been identified sufficiently such that the mean (μ_i) and variance (σ^2_i) are known. The laws of expectation provide that

$$\mu_T = \sum_{i=1}^{n} \mu_i \tag{1}$$

where μ_T refers to the mean of the total cost pdf (Neter et al. 1989). This is, the mean of the total cost pdf is equal to the sum of the means of the component element pdfs. Likewise, the variance of the total cost distribution is given by

$$\sigma^2_T = \sum_{i=1}^{n} \sigma^2_i + 2\sum_{i=1}^{n} \sum_{j=i-1}^{n} \rho_{ij}\sigma_i\sigma_j \tag{2}$$

where σ^2_T refers to the variance of the total cost pdf and σ_i refers to the standard deviation of cost element i (Neter et al. 1989). The variance is affected by the degree of correlation between cost elements. The correlation coefficients ρ_{ij} ranges from -1 to +1 indicating the strength of the relationship between cost element i and j. Values close to 0 indicate weak correlation, while values close to -1/+1 indicate a strong relationship. The sign indicates direction of the correlation. A positive ρ_{ij} value is stating that cost elements i and j tend to move in the same direction if the correlation coefficient is positive (when element i increases, cost element j also tends to increase). A negative correlation coefficient implies the cost elements move in opposing directions.

If the cost elements are statistically independent, all ρ_{ij} are equal to 0 and the second term on the right hand side of (2) equals 0. As such, the variance of the total cost pdf is equal to the sum of the variances of the component element pdfs if cost

elements are independent. If some or all of the n cost elements are linearly dependent, the variance will be affected. If there is positive correlation, the variance of total cost will be larger than when the cost elements are independent. If there is negative correlation, it is possible for the variance of total cost to be smaller than when the cost elements are independent.

Assuming independent cost elements in a cost uncertainty analysis is stating a belief that the cost of each and every element in the system is unrelated to the cost of all other elements. For two cost elements to be statistically independent, the analyst must believe that changes in the cost of element A are unconnected to changes in the cost element B. In other words, if the cost of element A turns out higher that expected, this in no way provides an indication that the cost of element B would be either higher or lower than expected. For many systems, the assumption of independent cost elements is difficult to defend. There is often ample reason to suspect that cost elements are correlated[3]. Garvey offers some evidence to support this belief (Garvey 1990). Cost data was gathered from electronic systems at level 2 of the WBS. Analysis showed that strong positive correlation existed between prime mission product costs and support costs.

In addition to the correlation affecting the shape of the total cost pdf, the shape of the cost element distributions will have an impact. Cost element distributions can be developed either from data or by applying subjective judgments. Cost element distributions are almost exclusively unimodal–that is, the pdf has a single high point (mode). An important characteristic of the cost element pdf is the skewness of the distribution. Skewness is simply a measure of the asymmetry of the distribution (see Figure 2). Positively skewed distributions can be used to represent cost elements that have a risk of a significant cost overrun. A positively skewed distribution is one in which the most likely cost is closer to the low cost value than to the high cost value. In addition, a positively skewed distribution is appropriate if there is a higher probability of

[3] Because pdfs used in this way usually represent subjective probabilities, it is difficult to gather data to confirm these suspicions. Garvey (1990) presents data from electronic systems that indicates level 2 WBS elements are positively correlated.

overrunning the most likely cost than underrunning that cost. This is because positively skewed distribution has more area under the curve (probability) to the right of the mode (most likely) than to the left of the mode.

Insert Figure 2 Here

If all of the cost elements are symmetric, it would be reasonable to expect that the sum of the distributions (the total cost pdf) will also be symmetric. This might also be a reasonable expectation if there were roughly an equal number of positively skewed distributions as negatively skewed distributions. If all cost elements were skewed in one direction, it might be reasonable to expect that the sum of the distributions to begin to skew in the same direction. It would seem reasonable that increasing correlation would reinforce and amplify this tendency.

3. The Approach

This study examined two pdfs; namely, the normal and the lognormal distributions. The normal distribution is used in heuristics proposed by both Kazanowski (1983) and Garvey (1990). Kazanowski recommended the normal pdf without offering support for his recommendation. Garvey selected the normal distribution based on analysis of cost data gathered from projects that were software development intensive (Garvey 1990). The lognormal distribution was originally proposed by Young (1992). The motivation for proposing the lognormal pdf stemmed from the author's cumulative experience accomplishing Monte Carlo simulation-based uncertainty analyses. He had observed that the total cost pdf tended to be positively skewed, and that the skew increased as the degree of (positive) correlation increased. The lognormal distribution is a positively skewed distribution in which the skewness increases as the variance

increases (Book 1994). The link between correlation and variance contained in (2) offers support for this line of reasoning.

A simple experiment was conducted to characterize the accuracy of the heuristic methods. The experiment involved executing a Monte Carlo simulation of an additive cost uncertainty model. A triangular distribution with endpoints of 0 and 100 was selected as the pdf for each cost element. For each experimental condition, 2,000 iterations of the Monte Carlo simulation were conducted. Crystal Ball™ was used to conduct the simulations. The accuracy of the distributions were examined under varying conditions. Specifically, the number of cost elements, the strength of cost element correlation, and the skewness of the cost element pdfs were systematically varied.

The number of cost elements was varied to determine if the size of the estimate affects heuristic accuracy. The number of cost elements was set at either 10 or 25. These values were selected because 10 represents a reasonable lower bound on the number of cost elements in a real cost estimate, while 25 is characteristic of somewhat larger projects.

The overall correlation level between cost elements was characterized as either weak, moderate, or strong. This study considered only positive correlation among cost elements. Following Garvey's suggested guidelines, individual ρ_{ij}s values were set to either 0, .25, .50, .75 or 1.0[4]. The degree of correlation was controlled by adjusting the proportion of ρ_{ij} values that assumed each value. The target proportions used to achieve the correlation settings are identified in Table 1. Within each setting, correlation coefficients are assigned to cost element pairs at random[5].

[4] Crystal Ball™ interprets the ρ_{ij} values as Spearman's correlation coefficients which are based on ranks. The coefficients found in (2) refer to the more common Pearson's correlation coefficient. Burgess computed some baseline cases using both interpretations and concluded that "... the difference in output was not discernable" (Burgess and Book 1993, pg. 4).

[5] Assigning correlation coefficients at random created the possibility that the resulting correlation matrix would be inconsistent (less than positive semidefinite). In this event, the correlation values were adjusted using the procedure available in Crystal Ball™. The algorithm by Lurie (1993) is capable of adjusting the matrix in the event Crystal Ball™ is not available.

Insert Table 1 Here

The aggregate skewness of the cost element pdfs was defined as moderate and high. As a first step, the individual cost element pdfs were established as either skewed or symmetrical. A symmetrical distribution was defined by randomly selecting a mode between the values of 35 and 60. A skewed cost element pdf was created by selecting a mode at random between the values of 0 and 35. Aggregate skewness was set by controlling the proportion of skewed and symmetrical distributions included in the simulation run. The moderate skewness setting corresponded to 80% symmetrical pdfs and 20% skewed pdfs. High skewness was established by reversing these proportions. To achieve high skewness, 80% of the cost element distributions were skewed distributions and 20% were symmetrical.

Insert Figure 3 Here

Each combination of the 3 experimental factors was examined, consequently 12 (2x3x2) simulation runs were accomplished. For illustration, the inputs for one of the simulation run (low settings for all factors) are contained in Figure 3. Each of the 12 resulting total cost pdfs was compared to the normal and lognormal distributions with mean and variance established by (1) and (2). The Chi-square goodness of fit test was selected to test the null hypothesis that the pdf generated by the simulation was indistinguishable from the hypothesized distribution (either normal or lognormal). The Chi-square statistic was calculated using 100 equiprobable intervals. The significance value for the test was fixed at $1-\alpha=.90$. The significance level was not set higher in order to maintain the power of the test and protect against the probability of making a type II error (falsely accepting the null hypothesis) (Law and Kelton 1991).

4. Results

The results of simulating the inputs contained in Figure 3 are provided in Figure 4. The total cost pdf has been converted to a cdf for illustration purposes. For this combination of factor settings, the normal cdf is almost identical to the simulated cdf. The lognormal cdf departs slightly from the simulation. The Chi-square results are consistent with Figure 4. At an $\alpha = .1$, the test indicates that the simulation results and the hypothetical normal pdf are equivalent distributions. At the same level of significance, however, the hypothesis that the simulated total cost pdf is equivalent to lognormal pdf is rejected. For this condition (10 cost elements that are weakly correlated and moderately skewed), a heuristic based on the normal distribution provides the same total cost pdf as a Monte Carlo simulation.

Insert Figure 4 Here

The calculated Chi-square statistics for all 12 setting combinations are provided in Tables 2 and 3. If the Chi-square statistic exceeds 117.4 ($\chi^2_{99,.9}$), the null hypothesis stating the distributions are equivalent is rejected. For the normal distribution, six of the 12 experimental conditions passed the goodness of fit test. All simulations with weak correlation generated pdfs that were indistinguishable from the normal pdf. In contrast, all conditions with strong correlation failed the goodness of fit test. For the moderately correlated cases, the goodness of fit results were influenced by the skewness. Given moderate correlation, the moderately skewed cases passed the test while the highly skewed cases failed. The number of cost elements did not affect the results of the Chi-square tests. For the lognormal, all conditions generated total cost pdfs that failed the goodness of fit test. The simulation pdfs were statistically different from the theoretical lognormal pdf.

Insert Tables 2 and 3 Here

There are a couple of points to note regarding the results from the Chi-square goodness of fit tests. The first point is that results from the normal distribution were much better than those from the lognormal distribution. None of the lognormal Chi-square statistics were close to acceptance. On the other hand, six conditions passed the test for the normal distribution while one case was very close to passing. The calculated Chi-square values for the 25 element, high correlation, moderate skewness case would have passed the goodness of fit test if alpha were reduced (p values for this case is approximately .04. As mentioned previously, however, significance can only be increased at the expense of the power of the test. To do so in this case would increase the risk of wrongly concluding that the distributions are equivalent.

A second insight can be gained by closer examination of the cases that fail to pass the goodness of fit test. The insight is gained by determining the location of the departures of the heuristic and the simulation results. Consider the case of 25 elements with strong correlation and high skewness. Judging by Chi-Squared statistic, the heuristics had the most trouble matching this case. Figure 5 shows the heuristic and simulated pdfs, while Figure 6 gives the cumulative data and cdfs for the same case. Note that the greastest departures between the simulated and heuristic distributions occur in the tails of the distribution. For this case, the departures are most dramatic in the left tail of the distribution. This is characteristic of all cases that failed the goodness of fit test. This is important because departures in this area are not particularly important in cost uncertainty analyses. When a total cost cdf is used in decision making, it is highly unlikely that the focus will be directed toward cost levels that provide a 95% probability of overrunning.

Insert Figures 5 and 6 Here

If the total cost pdf is used to establish or validate funding levels, the range of interest will most likely be towards the center of the distribution, perhaps between the 40% and 90% levels. A decision to fund a program below the 40% level is a decision to accept a much higher probability of overrun than underrun. A decision to fund at a level in excess of 90% is probably a luxury that cannot currently be considered.

Tables 4 through 7 indicate the quality of the heuristic in this central region of the distribution. The tables present percentage departure of the heuristic pdf from the simulation results. A small percentage is evidence that the heuristic is providing a funding level that is very close to the level provided by the simulation. For example, consider the failed case of 25 cost elements that are strongly correlated and highly skewed. Suppose a decision maker is considering funding the program and is willing to accept a 30% probability of an overrun. If Monte Carlo simulation had been used to generate the total cost cdf, the answer would be a funding level of $1,187.83. If the normal pdf had been used as a heuristic in place of the simulation, the funding level would be $1,204.46. On a percentage basis, the heuristic departs from the simulation results by +1.4%.

Insert Tables 4 through 7 Here

The percentage departures show that the heuristics provide results that are very close to the simulation in the center of the distribution. As expected, the experimental conditions that passed the goodness of fit tests provide very small percentage departures. Of particular interest are the departures for the conditions that failed the goodness of fit tests. Many of these percentage departures are quite small. For example, the 10

element, moderate skewness, high correlation condition failed the goodness of fit test for the normal distribution, but the percentage departures in the center range are comparable to (and in some cases smaller than) conditions that passed the test. The percentage departures for the lognormal distribution are greater than for the normal distribution in almost every case. Furthermore, the departures for the normal distribution tend to be positive while the lognormal departures tend to be negative. The normal distribution can therefore be characterized as conservative in that when departures occur, they tend to be overestimations rather than underestimations. The opposite can be said of the lognormal distribution.

Insert Table 8 Here

Finally, a summary of the computer run times is provided in Table 8. The simulations were run on a Zentih 486/33 PC equipped with 8MB RAM. Microsoft Excel™ 4.0 was used to host Crystal Ball™ version 3.0. The times reported are the execution time only for 2,000 iterations of the simulation. Increasing the number of iterations will cause a linear increase in run time. Depending on the number of simulations open at one time, input (reading files and recalling stored runs) and output (writing files and creating reports) can increase the time required by 5 to 10 minutes.

5. Conclusions

Heuristics for cost uncertainty analysis have been in existence for years, yet their accuracy has not been investigated. This research is a first step toward establishing the suitability of the normal and lognormal pdfs as tools to create the total cost pdf for an additive uncertainty model.

While some of these results were anticipated, this study supports some surprising conclusions. We expected that the fit of the normal distribution would deteriorate as

positive correlation and positive skewness increased. The surprise came in the closeness of the fit provided by the normal distribution. The normal distribution provides an excellent approximation for the simulated distribution. For cases of weak and moderate correlation and moderate skewness, the approximation is almost exact. For more severe conditions, the approximation remains good. Referring back to Tables 4 and 5, the worst departure for the normal distribution between the 40% and 90% levels is 5.3%. Most departures are much smaller than this. In this central region of the distribution, the difference between the simulated and the heuristic total cost pdfs is less than 2% in the vast preponderance (85%) of the cases.

The lognormal distribution did not prove to be as accurate as the normal distribution. In almost all cases, the normal distribution provided a better approximation of the simulation results than the lognormal. The lognormal fit degraded slightly as the degree of correlation and skewness increased. This result was somewhat unexpected as it was anticipated that these conditions would generate distributions more similar to the lognormal.

The implications of this study are important. If the model used in a cost uncertainty analysis resembles the models used in this research, an analyst can have confidence in applying a heuristic method. Under a wide variety of conditions, the normal distribution can be used to generate the total cost pdf without sacrificing the accuracy of a Monte Carlo simulation.

A cost analyst faced with a cost uncertainty analysis can use the results of this study in making the trade-off between the normal heuristic and Monte Carlo simulation. To do so, the analyst should characterize the correlation in their analysis as weak, moderate or strong, and the skewness as moderate or high. To characterize the correlation, compare the correlation matrix to those in Table 1. Calculating the average correlation coefficient may help in this comparison. To characterize the skewness, count the number of cost element distributions that are skewed (i.e. they have a mode in the lower one third of the range). As a rough guideline, the analyst might choose a 50% cutoff. If fewer than 50% of the elements are skewed, use the moderate skewness results as the guideline. If there are greater than 50% skewed distributions, the high skewness

results should be considered. Once the correlation and skewness of the uncertainty analysis are characterized, the goodness of fit results and the percentage departures provide an estimate of the accuracy that might be sacrificed if the normal distribution is used in place of a simulation.

To complete the trade-off analysis, Table 8 can be used to estimate the time required to run a Monte Carlo simulation. The time required for calculating the normal pdf is negligible. Most spreadsheets provide built-in functions that accomplish the calculations. Microsoft Excel™ has two such functions. If the analyst supplies the mean and variance of the normal distribution (provided by (1) and (2)), one function supplies the cost for a percentage of interest, while the other function will provide the percentage for a given cost of interest. If the spreadsheet capabilities are not available, the normal tables contained in any statistics book can be used.

If the both time and accuracy are of a premium, a combined strategy might be attractive. The heuristic approach could be used while the uncertainty analysis is in process to provide interim results. Monte Carlo simulation could then be applied to determine the final total cost pdf. This strategy would provide a significant time savings over the course of the uncertainty analysis. This time savings would be gained with confidence that the final total cost pdf will not differ significantly from the interim results provided by the heuristic.

There are certain conditions under which the results provided here should **not** be used as a guideline. These results only apply if an additive uncertainty model is being used. With an additive model, equations (1) and (2) can be used. If the model is multiplicative, there is no reliable method for determining the mean and variance of the total cost pdf outside of simulation.

A second limitation of this study is the assumption of equally weighted cost element pdfs. Each of the pdfs in this study ranged from 0 to 100. As a result, the variance of each distribution was roughly equal. It is possible in real cost uncertainty analyses to have a handful of cost elements with larger variances than the remainder of the cost elements. When this occurs, the distributions with the large cost variances dominate the others and have a much stronger influence in determining the shape of the

total cost pdf. The results of this study are applicable as long as there are close to 10 (or more) dominant cost elements. If there are fewer than 10, the recommendation is that Monte Carlo simulation be applied.

Finally, the use of the triangular distribution may be viewed by some as a limitation of this study. The primary reason for using triangular distributions for the cost elements is because it is a fairly common practice to apply the triangular distribution in the absence of data. The parameters of the distribution (low, most likely and high) correspond directly to the subjective assessments commonly asked of technical experts. Triangular distributions were also used to ensure the study was conservative. The triangular distribution is conservative in the sense that it posses more area in the tails of the distribution than do the distributions with tapered tails (i.e. the beta, lognormal, gamma or normal). As such, use of the triangular distribution is likely to increase the size of the tails of the total cost pdf over that which would result from using tapered distributions. It seems reasonable to speculate that if tapered distributions were applied in place of the triangular distributions, the tails of the total cost pdf would shrink and the fit of the normal and lognormal distributions would improve. The use of the triangular distribution is the most likely reason that the fit of the lognormal distribution degraded as correlation and skewness increased.

This study has examined the suitability of the normal and the lognormal distributions for use in a cost uncertainty analysis. This paper provides guidelines that analysts can use to effectively employ heuristic methods. Hopefully, these guidelines will prove useful to cost analysts confronted with the challenge of performing a cost uncertainty analysis.

Table 1 - Correlation Settings

ρ_{ij}	Weak	Moderate	Strong
0	50%	17%	—
.25	50%	33%	17%
.50	—	33%	33%
.75	—	17%	33%
1.0	—	—	17%
Mean ρ_{ij} value	.125	.375	.625

Table 2 - Results of Chi-Square Goodness of Fit Tests*
Simulated vs. Normal Distribution

		10 Elements Skewness		25 Elements Skewness	
		Moderate	High	Moderate	High
Correlation	Weak	68.1	101.5	91.3	98.7
	Moderate	109.8	145.6	114.3	143.1
	Strong	134.2	199.7	125.9	193.6

* Shaded regions indicate conditions for which Ho is rejected at alpha = .1

Table 3 - Results of Chi-Square Goodness of Fit Tests*
Simulated vs. Lognormal Distribution

		10 Elements Skewness		25 Elements Skewness	
		Moderate	High	Moderate	High
Correlation	Weak	171.6	166.7	134.5	134.4
	Moderate	329.1	264.8	236.8	225.9
	Strong	401.9	499.9	343.6	310.5

* Shaded regions indicate conditions for which Ho is rejected at alpha = .1

Table 4 - Percentage Departure of Heuristic from Simulated*

Normal Distribution/10 Elements**

		Moderate Skewness			High Skewness		
	Correlation	Weak	Moderate	Strong	Weak	Moderate	Strong
	40%	0.3%	1.6%	2.0%	1.5%	3.4%	5.3%
	50%	0.4%	1.0%	0.6%	0.7%	2.3%	4.6%
Probability of	60%	0.3%	-0.3%	0.3%	0.9%	1.3%	3.6%
Not Overrunning	70%	0.3%	-0.6%	-0.3%	0.4%	1.3%	0.2%
	80%	0.1%	-0.8%	-0.8%	-0.1%	-0.3%	-0.8%
	90%	-0.4%	-0.3%	-0.7%	-0.5%	-0.5%	-2.4%

* (Normal-Simulated)/Simulated

** Shaded regions indicate conditions for which failed goodness of fit test

Table 5 - Percentage Departure of Heuristic from Simulated*

Normal Distribution/25 Elements**

		Moderate Skewness			High Skewness		
	Correlation	Weak	Moderate	Strong	Weak	Moderate	Strong
	40%	0.6%	0.5%	1.5%	0.5%	1.6%	4.6%
	50%	0.1%	0.9%	0.7%	0.1%	1.6%	4.3%
Probability of	60%	-0.2%	0.0%	0.2%	0.1%	1.3%	3.3%
Not Overrunning	70%	-0.1%	0.3%	-0.5%	0.7%	0.6%	1.4%
	80%	0.3%	0.7%	-0.3%	0.7%	0.7%	0.1%
	90%	-0.4%	0.4%	1.2%	0.2%	0.0%	-1.2%

* (Normal-Simulated)/Simulated

** Shaded regions indicate conditions for which failed goodness of fit test

Table 6 - Percentage Departure of Heuristic from Simulated*

Lognormal Distribution/10 Elements**

		Moderate Skewness			High Skewness		
	Correlation	Weak	Moderate	Strong	Weak	Moderate	Strong
Probability of Not Overrunning	40%	-1.5%	-2.0%	-3.3%	-1.0%	-1.9%	-3.1%
	50%	-1.6%	-3.0%	-5.3%	-2.0%	-3.7%	-5.0%
	60%	-1.6%	-4.1%	-5.4%	-1.8%	-4.5%	-5.8%
	70%	-1.2%	-3.9%	-5.2%	-1.8%	-3.7%	-7.8%
	80%	-0.7%	-2.7%	-3.9%	-1.3%	-3.4%	-6.1%
	90%	0.2%	0.5%	0.0%	0.2%	0.2%	-2.2%

* (Lognormal-Simulated)/Simulated

** Shaded regions indicate conditions for which failed goodness of fit test

Table 7 - Percentage Departure of Heuristic from Simulated*

Lognormal Distribution/25 Elements**

		Moderate Skewness			High Skewness		
	Correlation	Weak	Moderate	Strong	Weak	Moderate	Strong
Probability of Not Overrunning	40%	-0.8%	-2.6%	-3.5%	-1.6%	-2.8%	-3.3%
	50%	-1.5%	-2.5%	-4.9%	-2.2%	-3.3%	-4.7%
	60%	-1.7%	-3.2%	-5.3%	-2.2%	-3.5%	-5.6%
	70%	-1.4%	-2.4%	-5.1%	-1.2%	-3.4%	-6.1%
	80%	-0.4%	-0.9%	-3.2%	-0.3%	-1.8%	-4.8%
	90%	0.1%	1.2%	1.9%	0.8%	0.8%	-0.9%

* (Lognormal-Simulated)/Simulated

** Shaded regions indicate conditions for which failed goodness of fit test

Table 8 - Elapsed Time for Monte Carlo Simulation Runs*

	10 Cost Elements		25 Cost Elements	
Correlation	Moderate Skewness	High Skewness	Moderate Skewness	High Skewness
Weak	4:16	4:22	14:34	12:21
Moderate	5:08	4:21	9:20	8:57
Strong	4:25	4:21	7:23	9:28

* Zenith 486/33 running Crystall Ball™ ver 3.0 for 2,000 iterations. Reported time is minutes and seconds for execution only: input/output time excluded.

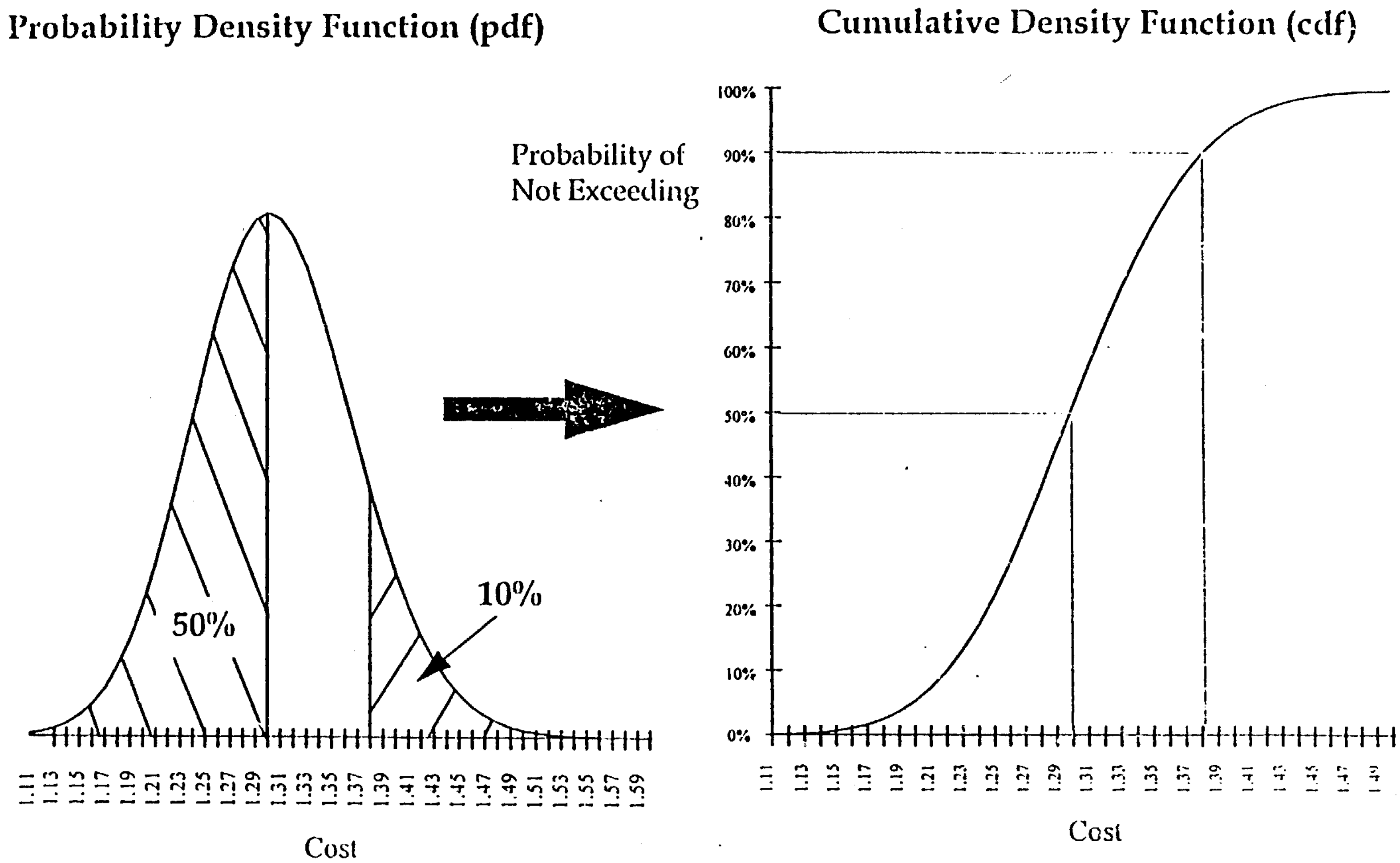

Figure 1 - The Total Cost Probability Density Function (pdf) and Cumulative Density Function (cdf)

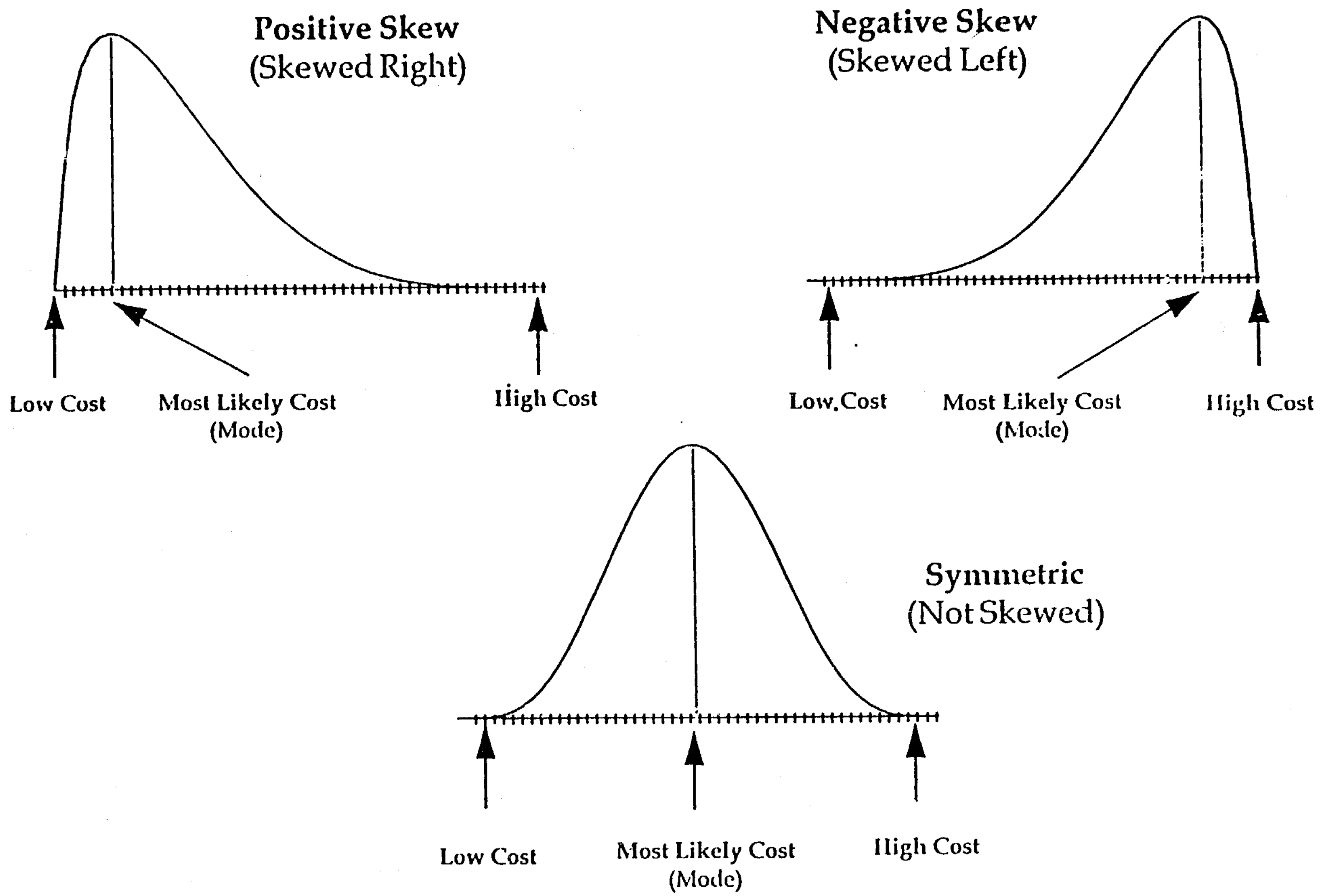

Figure 2 - PDF Skewness

Number of Cost Elements:	10
Degree of Correlation:	Weak
Degree of Skewness:	Moderate

All cost elements distributed Triangular

Cost Element	Low	Most Likely	High
1	0	12.1	100
2	0	46.9	100
3	0	57.8	100
4	0	22.7	100
5	0	56.1	100
6	0	46.0	100
7	0	54.2	100
8	0	42.2	100
9	0	38.6	100
10	0	37.3	100

Total Cost	
Mean	471
Var	8,854
Std Dev	94

Correlation Matrix

	1	2	3	4	5	6	7	8	9	10
1	1									
2	0	1								
3	0	0.25	1							
4	0	0	0.25	1						
5	0.25	0	0.25	0.25	1					
6	0	0	0.25	0.25	0.25	1				
7	0.25	0	0	0	0	0.25	1			
8	0.25	0.25	0	0.25	0	0	0	1		
9	0.25	0.25	0	0.25	0.25	0.25	0.25	0	1	
10	0	0.25	0	0.25	0	0	0	0	0	1

Figure 3 - Inputs for 10 element, Weak Correlation, Moderate Skewness Setting

Cumulative Probability	Simulated	Normal	% Departure Normal	Lognormal	%Departure Lognormal
2.5%	$289	$287	-0.8%	$314	8.5%
5.0%	$319	$316	-0.6%	$334	4.8%
10.0%	$355	$351	-1.3%	$359	1.0%
20.0%	$391	$392	0.3%	$391	0.1%
30.0%	$420	$422	0.4%	$417	-0.9%
40.0%	$446	$447	0.3%	$440	-1.5%
50.0%	$470	$471	0.4%	$462	-1.6%
60.0%	$494	$495	0.3%	$486	-1.6%
70.0%	$519	$521	0.3%	$513	-1.2%
80.0%	$550	$550	0.1%	$546	-0.7%
90.0%	$595	$592	-0.4%	$595	0.2%
95.0%	$628	$626	-0.3%	$640	1.9%
97.5%	$658	$656	-0.3%	$681	3.5%

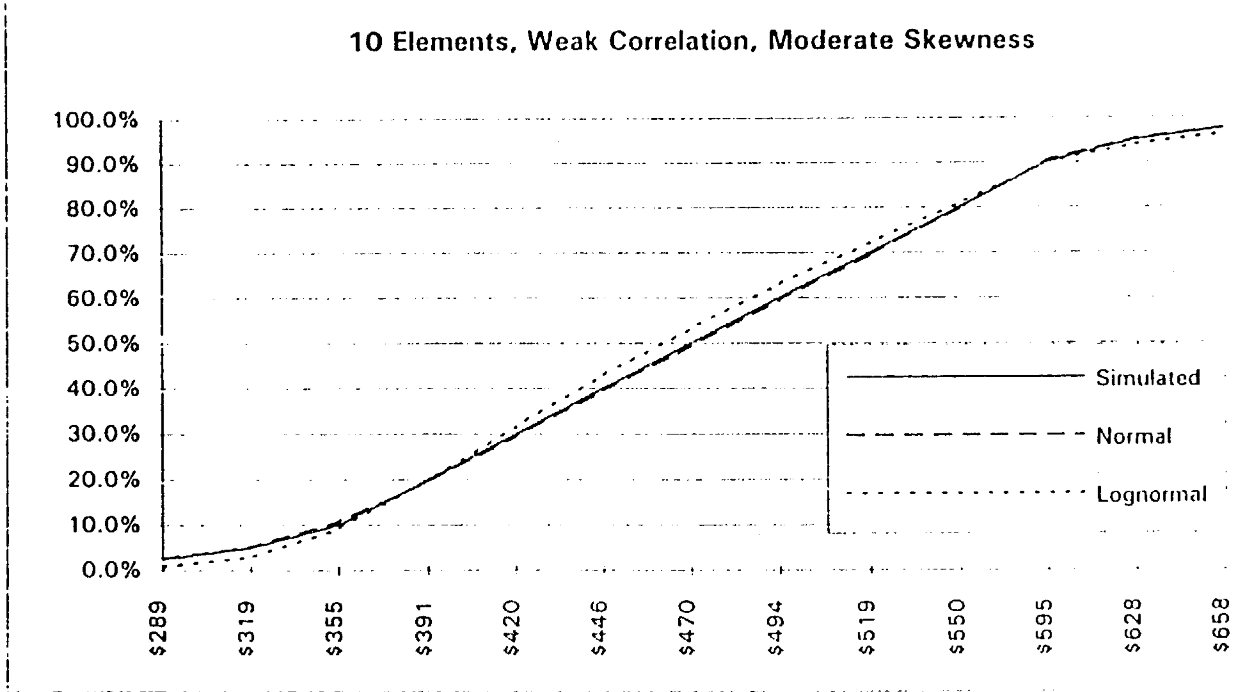

Figure 4 - Results for 10 Elements, Weak Correlation, Moderate Skewness Setting

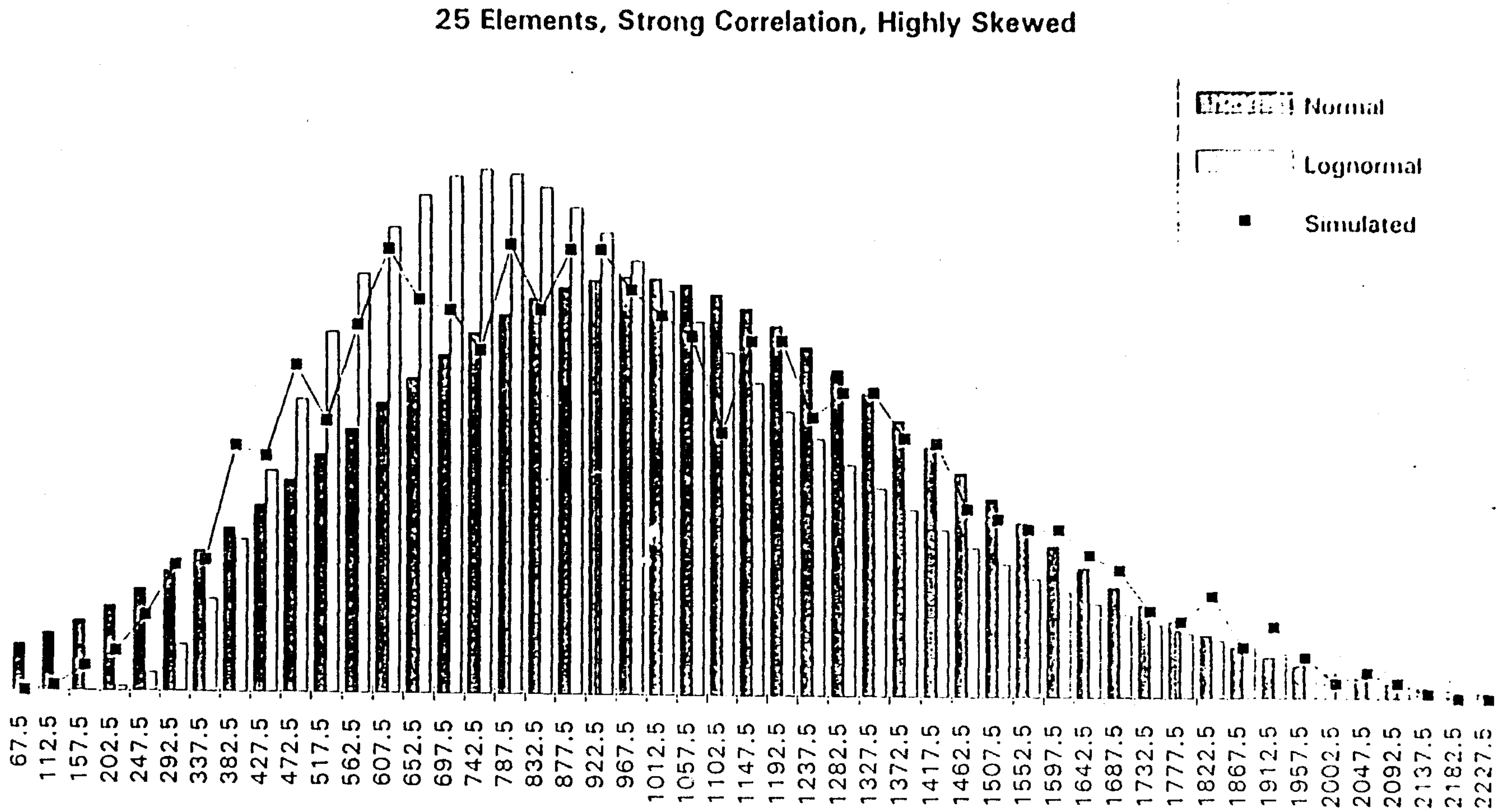

Figure 5 - Total Cost PDF for 25 Elements, Strong Correlation and High Skewness

Cumulative Probability	Simulated	Normal	% Departure Normal	Lognormal	% Departure Lognormal
2.5%	$311	$126	-59.6%	$388	24.9%
5.0%	$379	$262	-30.7%	$444	17.2%
10.0%	$467	$420	-10.1%	$518	10.8%
20.0%	$602	$611	1.6%	$624	3.8%
30.0%	$713	$749	5.1%	$714	0.2%
40.0%	$828	$867	4.6%	$801	-3.3%
50.0%	$937	$977	4.3%	$893	-4.7%
60.0%	$1,052	$1,087	3.3%	$994	-5.6%
70.0%	$1,188	$1,204	1.4%	$1,115	-6.1%
80.0%	$1,341	$1,342	0.1%	$1,276	-4.8%
90.0%	$1,553	$1,533	-1.2%	$1,538	-0.9%
95.0%	$1,702	$1,691	-0.7%	$1,795	5.4%
97.5%	$1,830	$1,828	-0.1%	$2,052	12.1%

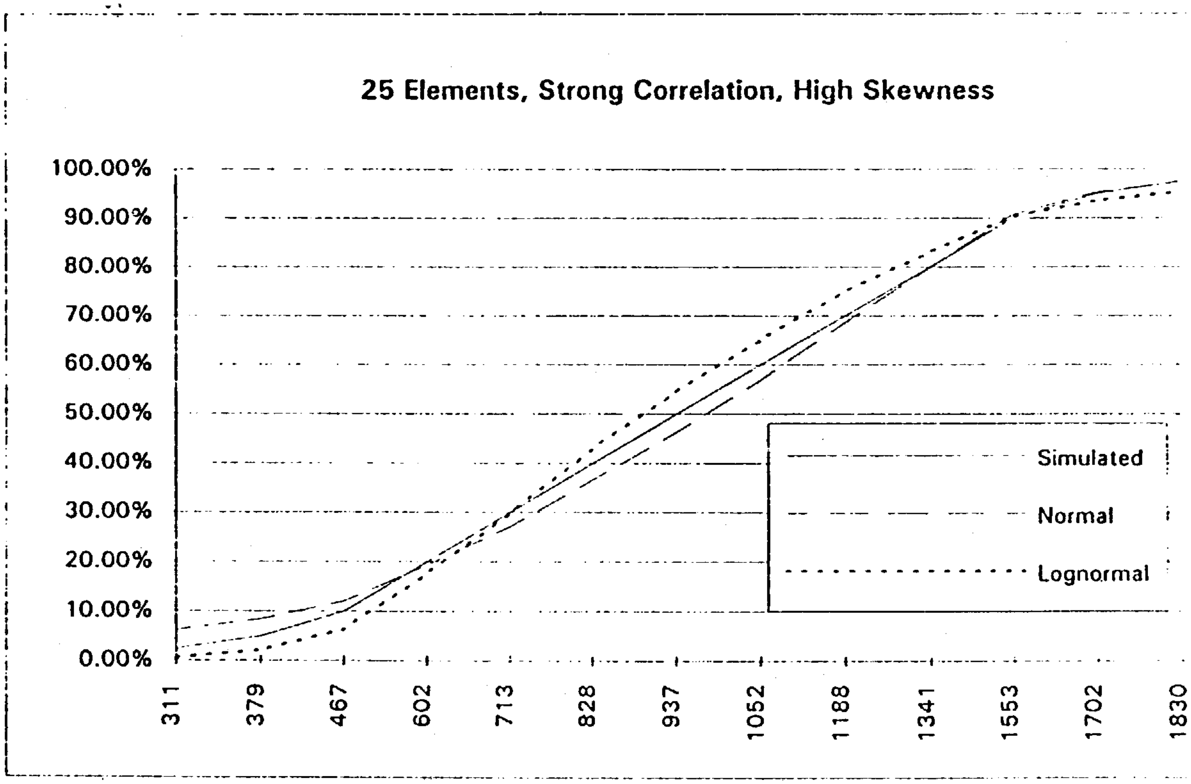

Figure 6 - Results for 25 Elements, Strong Correlation, High Slewness Setting

REFERENCES

Book, S.A., "Developments in Cost Risk Analysis", lecture to QMGT 680 - Project Risk Analysis, Air Force Institute of Technology, 27 Jan 1994.

Burgess, E.L. and S.A. Book, "Validation Report on Crystal Ball™ Risk Modeling Software", Technical Report of the Resources Analysis Department, The Aerospace Corporation, 5 Jan 1993 (access restricted to U.S government only)

Deinemann, P.F., "Estimating Cost Uncertainty Using Monte Carlo Techniques", The Rand Corporation, Report RM-4854-PR, January, 1966.

Garvey, P.R., "A General Analytic Approach to System Cost Uncertainty Analysis", Cost Analysis and Estimating: Tools and Techniques, ed by Greer and Nussbaum, Springer Verlag, 1990, pp 161-181.

Graham, D. "Cost and Schedule Risk Analysis for the 1992 Independent Cost Estimate of the Milstar Satellite Program", paper presented to the 26th Annual DoD Cost Analysis Symposium, Sept 1992.

Iman, D.L. and W.J. Conover, "A Distribution-Free Approach to Inducing Rank Correlation among Input Variables", Communications of Statist.-Simula. Computa., Vol 11, No. 3, 1982, pp 311-334.

Kazanowski, A.D., "A Quantitative Methodology for Estimating Total System Cost Risk". The Aerospace Corporation, Technical Report # SD-TR-83-44, 3 Jun 1983.

Law, A.M., and W.D. Kelton, Simulation Modeling and Analysis, 2nd Ed., McGraw Hill Inc, New York, 1991.

Lurie, P.M., "Simulating Correlated Distributions with Bounded Domains", The Institute for Defense Analysis, Technical Report IDA-P-2732, Sept 1992.

Neter, J., W. Wasserman and M.H. Kutner, Applied Linear Regression Models, 2nd Ed., Richard D. Irwin Inc, Homewood IL, 1989.

Young, P. H., "FRISK - Formal Risk Assessment of System Cost Estimates", proceedings of the 1992 Aerospace Design Conference, The American Institute of Aeronautics and Astronautics, February 1992.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
26 Aug 94

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
An Investigation of the Accuracy of Heuristic Methods for Cost Uncertainty Analysis

5. FUNDING NUMBERS

6. AUTHOR(S)
Major Wendell P. Simpson III, Ph.D.
Major Kevin P. Grant, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute Of Technology
Graduate School of Logistics and Acquisition Management
2950 P Street
Wright-Patterson AFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB, OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AU-AFIT/LA-TR-94-1

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Cost uncertainty analysis has received a great deal of attention over the last several years. The purpose of a cost uncertainty analysis is to identify the cost and schedule implications associated with program uncertainties. Common methods for uncertainty analysis characterize the possible cost and schedule outcomes of a project using a probability density function (pdf). Heuristic methods have been proposed for uncertainty analysis that assume the shape of the total cost pdf is either normally or lognormally distributed. While experienced analysts feel these distributions provide reasonable approximations, little evidence exists to either confirm or refute these presumptions. An experiment is conducted in which number of cost elements, the degree of skewness of the cost elements, and the degree of correlation between cost elements are varied systematically. The resulting total cost pdfs are compared to the heuristic distributions using goodness of fit tests. The results show that the normal distribution provides an excellent approximation for the simulateddistributions. Guidelines are offered that help the cost analyst determine whether these heuristics ought to be applied in a cost uncertainty analysis.

14. SUBJECT TERMS
Cost Analysis, Cost Uncertainty Analysis

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	